

University of Groningen

A Generalized Characterization of Algorithmic Probability

Sterkenburg, Tom F.

Published in:
Theory of computing systems

DOI:
[10.1007/s00224-017-9774-9](https://doi.org/10.1007/s00224-017-9774-9)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):
Sterkenburg, T. F. (2017). A Generalized Characterization of Algorithmic Probability. *Theory of computing systems*, 61(4), 1337-1352. <https://doi.org/10.1007/s00224-017-9774-9>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

A Generalized Characterization of Algorithmic Probability

Tom F. Sterkenburg^{1,2}

Published online: 13 May 2017

© The Author(s) 2017. This article is an open access publication

Abstract An a priori semimeasure (also known as “algorithmic probability” or “the Solomonoff prior” in the context of inductive inference) is defined as the transformation, by a given universal monotone Turing machine, of the uniform measure on the infinite strings. It is shown in this paper that the class of a priori semimeasures can equivalently be defined as the class of transformations, by all compatible universal monotone Turing machines, of any continuous computable measure in place of the uniform measure. Some consideration is given to possible implications for the association of algorithmic probability with certain foundational principles of statistics.

Keywords Algorithmic probability · A priori semimeasure · Semicomputable semimeasures · Monotone turing machines · Principle of indifference · Occam’s razor

1 Introduction

Levin [23] first considered the transformation of the uniform measure λ on the infinite bit strings by a universal monotone machine U . This transformation λ_U is the function that for each finite bit string returns the probability that the string is gen-

This article is part of the Topical Collection on *Special Issue on Computability, Complexity and Randomness (CCR 2015)*

✉ Tom F. Sterkenburg
tom@cw.nl

¹ Algorithms and Complexity Group, Centrum Wiskunde & Informatica, Amsterdam, The Netherlands

² Faculty of Philosophy, University of Groningen, Groningen, The Netherlands

erated by machine U , when U is supplied a stream of uniformly random input (produced by tossing a fair coin, say). Levin attached to λ_U the interpretation of an “a priori probability” distribution, because λ_U dominates every other semicomputable semimeasure and so the initial assumption that a sequence is randomly generated from λ_U is in an exact sense the weakest of randomness assumptions.

Earlier on, Solomonoff [20] described in a somewhat less precise way a very similar definition. His motivation was an “a priori probability” distribution to serve as an objective starting point in inductive inference. In this context the definition is known under various headers, including “the Solomonoff prior” and “algorithmic probability”; and it has been associated with certain foundational principles from statistics, to explain or support its merits as an idealized inductive method.

As commonly presented, however, the association with two main such principles (firstly, the principle of *indifference*, and secondly, the principle of *Occam’s razor*) seems to essentially rest on the definition of λ_U as a universal transformation of the *uniform measure* λ .

This raises the question whether the *a priori semimeasures* (as we will call the functions λ_U here) must be defined, as they always are, as the universal transformations of the uniform measure, or that the *a priori semimeasures* can equivalently be defined as universal transformations of other computable measures.

The main result of this paper is that any *a priori semimeasure* can indeed be obtained as a universal transformation of *any* continuous computable measure. That is, for any continuous computable measure, an *a priori semimeasure* can equivalently be defined as giving the probabilities for finite strings being generated by a universal machine that is presented with a stream of bits sampled from *this* measure. More precisely, for any continuous computable measure μ , it is shown that the class of functions λ_U for all universal monotone machines U coincides with the class of functions μ_U (i.e., the transformation by U of μ) for all (μ -compatible) universal machines U .

This work will be done in Section 2. First, in the current section, we cover basic notions and notation (Section 1.1), discuss the characterization of the semicomputable semimeasures as the transformations via monotone machines of a continuous computable measure (Section 1.2), and the analogous characterization for semicomputable discrete semimeasures and prefix-free machines (Section 1.3).

1.1 Basic Notions and Notation

1.1.1 Bit Strings

Let $\mathbb{B} := \{0, 1\}$ denote the set of bits; \mathbb{B}^* the set of all finite bit strings; \mathbb{B}^n the set of bit strings σ of length $|\sigma| = n$; $\mathbb{B}^{\leq n}$ the set of bit strings σ of length $|\sigma| \leq n$; \mathbb{B}^ω the class of all infinite bit strings. The empty string is ϵ . The concatenation of bit strings σ and τ is written $\sigma\tau$; we write $\sigma \preceq \tau$ if σ is an *initial segment* of τ (so there is a ρ such that $\sigma\rho = \tau$; we write $\sigma < \tau$ if $\rho \neq \epsilon$). The initial segment of σ of length $n \leq |\sigma|$ is denoted $\sigma \upharpoonright_n$; the initial segment $\sigma \upharpoonright_{|\sigma|-1}$ is denoted σ^- . Strings σ and τ are *comparable*, $\sigma \sim \tau$, if $\sigma \preceq \tau$ or $\tau \preceq \sigma$; if σ and τ are not comparable we write $\sigma \nmid \tau$.

For given finite string σ , the class $\llbracket \sigma \rrbracket := \{\sigma X : X \in \mathbb{B}^\omega\} \subseteq \mathbb{B}^\omega$ is the class of infinite extensions of σ . Likewise, for $A \subseteq \mathbb{B}^*$, let $\llbracket A \rrbracket := \{\sigma X : \sigma \in A, X \in \mathbb{B}^\omega\}$.

1.1.2 Computable Measures

A probability measure over the infinite strings is generated by a *premeasure*, a function $m : \mathbb{B}^* \rightarrow [0, 1]$ that satisfies

1. $m(\epsilon) = 1$;
2. $m(\sigma 0) + m(\sigma 1) = m(\sigma)$ for all $\sigma \in \mathbb{B}^*$.

A premeasure m gives rise to an *outer measure* $\mu_m^* : \mathcal{P}(\mathbb{B}^\omega) \rightarrow [0, 1]$ by

$$\mu_m^*(\mathcal{A}) = \inf \left\{ \sum_{\sigma \in A} m(\sigma) : \mathcal{A} \subseteq \llbracket A \rrbracket \right\}.$$

By restricting μ_m^* to the measurable sets, i.e., the sets $\mathcal{A} \subseteq \mathbb{B}^\omega$ such that $\mu_m^*(\mathcal{B}) = \mu_m^*(\mathcal{B} \cap \mathcal{A}) + \mu_m^*(\mathcal{B} \setminus \mathcal{A})$ for all $\mathcal{B} \subseteq \mathbb{B}^\omega$, we finally obtain the corresponding (*probability*) *measure* μ_m , that satisfies $\mu_m(\llbracket \sigma \rrbracket) = m(\sigma)$ for all $\sigma \in \mathbb{B}^*$.

The *uniform (Lebesgue) measure* λ is given by the premeasure m with $m(\sigma) = 2^{-|\sigma|}$ for all $\sigma \in \mathbb{B}^*$. A measure μ is *nonatomic* or *continuous* if there is no $X \in \mathbb{B}^\omega$ with $\mu(\{X\}) > 0$.

We call a total real-valued function $f : \mathbb{B}^* \rightarrow \mathbb{R}$ *computable* if its values are uniformly computable reals: there is a computable $g : \mathbb{B}^* \times \mathbb{N} \rightarrow \mathbb{Q}$ such that $|g(\sigma, k) - f(\sigma)| < 2^{-k}$ for all σ, k . This allows us to talk about computable premeasures. A measure μ we then call *computable* if $\mu = \mu_m^*$ for a computable premeasure m .

1.1.3 Semicomputable Semimeasures

We call a total real-valued function $f : \mathbb{B}^* \rightarrow \mathbb{R}$ (*lower*) *semicomputable* if there are uniformly computable functions $f_t : \mathbb{B}^* \rightarrow \mathbb{Q}$ such that for all $\sigma \in \mathbb{B}^*$, we have $f_{t+1}(\sigma) \geq f_t(\sigma)$ for all $t \in \mathbb{N}$ and $\lim_{t \rightarrow \infty} f_t(\sigma) = f(\sigma)$.

Levin [23, Definition 3.6] introduced the notion of a semicomputable measure over the collection $\mathbb{B}^* \cup \mathbb{B}^\omega$ of finite and infinite strings. This is equivalent to a *semimeasure* over the infinite strings that is generated from a premeasure m that only needs to satisfy

1. $m(\epsilon) \leq 1$;
2. $m(\sigma 0) + m(\sigma 1) \leq m(\sigma)$ for all $\sigma \in \mathbb{B}^*$.

Following [5], we will simply treat a semimeasure as a function over the cones $\{\llbracket \sigma \rrbracket : \sigma \in \mathbb{B}^*\}$:

Definition 1 A *semicomputable semimeasure* is a function $\nu : \{\llbracket \sigma \rrbracket : \sigma \in \mathbb{B}^*\} \rightarrow [0, 1]$ such that $\nu(\llbracket \cdot \rrbracket) : \mathbb{B}^* \rightarrow [0, 1]$ is semicomputable, and

1. $\nu(\llbracket \epsilon \rrbracket) \leq 1$;
2. $\nu(\llbracket \sigma 0 \rrbracket) + \nu(\llbracket \sigma 1 \rrbracket) \leq \nu(\llbracket \sigma \rrbracket)$ for all $\sigma \in \mathbb{B}^*$.

Moreover, we follow the custom of writing $\nu(\sigma)$ for $\nu(\llbracket \sigma \rrbracket)$. Let \mathcal{M} denote the class of all semicomputable semimeasures.¹

1.2 Monotone Machines and Semicomputable Semimeasures

1.2.1 Machines

The following definition is due to Levin [10]. (Similar machine models were already described in [23], and by Solomonoff [20] and Schnorr [19]; see [3].)

Definition 2 A *monotone machine* is a c.e. set $M \subseteq \mathbb{B}^* \times \mathbb{B}^*$ of pairs of strings such that if $(\rho_1, \sigma_1), (\rho_2, \sigma_2) \in M$ and $\rho_1 \preceq \rho_2$ then $\sigma_1 \sim \sigma_2$.

We will not go into the concrete machine model that corresponds to the above abstract definition (see, for instance, [5, p. 145]); we only note that a machine M as defined above induces a function $N_M : \mathbb{B}^* \cup \mathbb{B}^\omega \rightarrow \mathbb{B}^* \cup \mathbb{B}^\omega$ by $N_M(X) = \sup_{\preceq} \{\sigma \in \mathbb{B}^* : \exists \rho \preceq X ((\rho, \sigma) \in M)\}$ (cf. [7]).

1.2.2 Transformations

Imagine that we feed a monotone machine M a stream of input that is generated from a computable measure μ . As a result, machine M produces a (finite or infinite) stream of output. The probabilities for the possible initial segments of the output stream are themselves given by a semicomputable semimeasure (as can easily be verified). We will call this semimeasure the *transformation* of μ by M .

Definition 3 The *transformation* μ_M of computable measure μ by monotone machine M is defined by

$$\mu_M(\sigma) := \mu(\llbracket \{\rho : \exists \sigma' \succ \sigma ((\rho, \sigma') \in M)\} \rrbracket).$$

1.2.3 Characterizations of \mathcal{M}

For every given semicomputable semimeasure ν , one can obtain a machine M that transforms the uniform measure λ to ν . Together with the straightforward converse that every function λ_M defines a semicomputable semimeasure, this gives a characterization of the class \mathcal{M} of semicomputable semimeasures as

$$\mathcal{M} = \{\lambda_M\}_M, \quad (1)$$

where $\{\lambda_M\}_M$ is the class of functions λ_M for all monotone machines M .

¹Semimeasures as defined here are often referred to as *continuous* semimeasures, in contradistinction to the *discrete* semimeasures defined in Section 1.3 below (cf. [5, 13]). Due to the possibility of confusion with the earlier meaning of “continuous” as synonymous to “nonatomic,” we will avoid this usage here.

A proof of this fact by the construction of an M that transforms λ to given ν was first outlined by Levin in [23, Theorem 3.2]. (Also see [13, Theorem 4.5.2].) Moreover, it can be deduced from [23, Theorem 3.1(b), 3.2] that \mathcal{M} can be characterized as the class of transformations of computable measures other than λ . Namely, we have that \mathcal{M} coincides with $\{\mu_M\}_M$ for any computable μ that is continuous.

A detailed construction to prove the characterization (1) was published by Day [4, Theorem 4(ii)]. (Also see [5, Theorem 3.16.2(ii)].) The following proof of the case for any continuous computable measure is an adaptation of this construction.

Theorem 1 (Levin) *For every continuous computable measure μ , and for every semicomputable semimeasure ν , there is a monotone machine M such that $\nu = \mu_M$.*

Proof Let ν be any semicomputable semimeasure, with uniformly computable approximation functions f_t . We construct in stages $s = \langle \sigma, t \rangle$ a monotone machine M that transforms μ into ν . Let $D_s(\sigma) := \{\rho \in \mathbb{B}^* : (\rho, \sigma) \in M_s\}$. \square

Construction Let $M_0 := \emptyset$.

At stage $s = \langle \sigma, t \rangle$, if $\mu(\llbracket D_{s-1}(\sigma) \rrbracket) = f_t(\sigma)$ then let $M_s := M_{s-1}$.

Otherwise, first consider the case $\sigma \neq \epsilon$. By Lemma 1 in [4] there is a set $R \subseteq \mathbb{B}^s$ of available strings of length s such that $\llbracket R \rrbracket = \llbracket D_{s-1}(\sigma^-) \rrbracket \setminus (\llbracket D_{s-1}(\sigma^-0) \rrbracket \cup \llbracket D_{s-1}(\sigma^-1) \rrbracket)$. Denote $x := \mu(\llbracket R \rrbracket)$, the amount of measure available for descriptions for σ , which equals $\mu(\llbracket D_{s-1}(\sigma^-) \rrbracket) - \mu(\llbracket D_{s-1}(\sigma^-0) \rrbracket) - \mu(\llbracket D_{s-1}(\sigma^-1) \rrbracket)$ because we ensure by construction that $\llbracket D_{s-1}(\sigma^-) \rrbracket \supseteq \llbracket D_{s-1}(\sigma^-0) \rrbracket \cup \llbracket D_{s-1}(\sigma^-1) \rrbracket$ and $\llbracket D_{s-1}(\sigma^-0) \rrbracket \cap \llbracket D_{s-1}(\sigma^-1) \rrbracket = \emptyset$. Denote $y := f_t(\sigma) - \mu(\llbracket D_{s-1}(\sigma) \rrbracket)$, the amount of measure the current descriptions fall short of the latest approximation of $\nu(\sigma)$. We collect in the auxiliary set A_s a number of available strings from R such that $\mu(\llbracket A_s \rrbracket)$ is maximal while still bounded by $\min\{x, y\}$.

If $\sigma = \epsilon$, then denote $y := f_t(\epsilon) - \mu(\llbracket D_{s-1}(\epsilon) \rrbracket)$. Collect in A_s a number of available strings from $R \subseteq \mathbb{B}^s$ with $\llbracket R \rrbracket = \mathbb{B}^\omega \setminus \llbracket D_{s-1}(\epsilon) \rrbracket$ such that $\mu(\llbracket A_s \rrbracket)$ is maximal but bounded by y .

Put $M_s := M_{s-1} \cup \{(\rho, \sigma) : \rho \in A_s\}$.

Verification The verification of the fact that M is a monotone machine is identical to that in [4].

It remains to prove that $\mu_M(\sigma) = \nu(\sigma)$ for all $\sigma \in \mathbb{B}^*$. Since by construction $\llbracket D_s(\sigma') \rrbracket \subseteq \llbracket D_s(\sigma) \rrbracket$ for any $\sigma' \succ \sigma$, we have that $\mu_{M_s}(\sigma) = \mu(\cup_{\sigma' \succ \sigma} \llbracket D_s(\sigma') \rrbracket) = \mu(\llbracket D_s(\sigma) \rrbracket)$. Hence $\mu_M(\sigma) = \lim_{s \rightarrow \infty} \mu(\llbracket D_s(\sigma) \rrbracket)$, and our objective is to show that $\lim_{s \rightarrow \infty} \mu(\llbracket D_s(\sigma) \rrbracket) = \nu(\sigma)$. To that end it suffices to demonstrate that for every $\delta > 0$ there is some stage s_0 where $\mu(\llbracket D_{s_0}(\sigma) \rrbracket) > \nu(\sigma) - \delta$. We prove this by induction.

For the base step, let $\sigma = \epsilon$. Choose positive $\delta' < \delta$. There will be a stage $s_0 = \langle \epsilon, t_0 \rangle$ where $f_{t_0}(\epsilon) > \nu(\epsilon) - \delta'$, and (since μ is continuous) $\mu(\llbracket \rho \rrbracket) \leq \delta - \delta'$ for all $\rho \in \mathbb{B}^{s_0}$. Then, if not already $\mu(\llbracket D_{s_0-1}(\epsilon) \rrbracket) > \nu(\epsilon) - \delta$, the latter guarantees that the construction will select a number of available strings in A_{s_0} such that $\nu(\epsilon) - \delta <$

$\mu(\llbracket D_{s_0-1}(\epsilon) \rrbracket) + \mu(\llbracket A_s \rrbracket) \leq f_{t_0}(\epsilon)$. It follows that $\mu(\llbracket D_{s_0}(\epsilon) \rrbracket) = \mu(\llbracket D_{s_0-1}(\epsilon) \rrbracket) + \mu(\llbracket A_s \rrbracket) > v(\epsilon) - \delta$ as required.

For the inductive step, let $\sigma \neq \epsilon$, and denote by σ' the one-bit extension of σ^- with $\sigma' \mid \sigma$. Choose positive $\delta' < \delta$. By induction hypothesis, there exists a stage s'_0 such that $\mu(\llbracket D_{s'_0}(\sigma^-) \rrbracket) > v(\sigma^-) - \delta'$. At this stage s'_0 , we have

$$\begin{aligned} \mu(\llbracket D_{s'_0}(\sigma^-) \rrbracket) - \mu(\llbracket D_{s'_0}(\sigma') \rrbracket) &\geq \mu(\llbracket D_{s'_0}(\sigma^-) \rrbracket) - v(\sigma') \\ &> v(\sigma^-) - \delta' - v(\sigma') \\ &\geq v(\sigma) - \delta', \end{aligned}$$

where the last inequality follows from the semimeasure property $v(\sigma^-) \geq v(\sigma) + v(\sigma')$. There will be a stage $s_0 = \langle \sigma, t_0 \rangle \geq s'_0$ with $f_{t_0}(\sigma) > v(\sigma) - \delta'$ and $\mu(\llbracket \rho \rrbracket) \leq \delta - \delta'$ for all $\rho \in \mathbb{B}^{s_0}$. Clearly, $\min\{\mu(\llbracket D_{s_0}(\sigma^-) \rrbracket) - \mu(\llbracket D_{s_0}(\sigma') \rrbracket), f_{t_0}(\sigma)\} > v(\sigma) - \delta'$. Then, as in the base case, if not already $\mu(\llbracket D_{s_0-1}(\sigma) \rrbracket) > v(\sigma) - \delta$, the construction selects a number of available descriptions such that $\mu(\llbracket D_{s_0}(\sigma) \rrbracket) > v(\sigma) - \delta$ as required.

Corollary 1 For every continuous computable measure μ ,

$$\{\mu_M\}_M = \mathcal{M}.$$

1.3 Prefix-Free Machines and Discrete Semimeasures

The notions of a semicomputable *discrete* semimeasure on the finite strings and a *prefix-free* machine can be traced back to Levin [11] and Gács [6], and independently Chaitin [1].

Definition 4 A *semicomputable discrete semimeasure* is a semicomputable function $P : \mathbb{B}^* \rightarrow \mathbb{R}^{\geq 0}$ such that $\sum_{\sigma \in \mathbb{B}^*} P(\sigma) \leq 1$.

Definition 5 A *prefix-free machine* is a partial computable function $T : \mathbb{B}^* \rightarrow \mathbb{B}^*$ with prefix-free domain.

Definition 6 The *transformation of computable measure μ by prefix-free machine T* is the semicomputable discrete semimeasure $Q_T^\mu : \mathbb{B}^* \rightarrow [0, 1]$ defined by

$$Q_T^\mu(\sigma) := \mu(\llbracket \{\rho : (\rho, \sigma) \in T\} \rrbracket).$$

Let \mathcal{P} denote the class of all semicomputable discrete semimeasures. Analogous to class \mathcal{M} and the monotone machines, class \mathcal{P} is characterized as the class of all prefix-free machine transformations of μ , for any continuous computable μ . The fact that every P can be obtained as a transformation of λ is usually inferred from the effective version of Kraft's inequality (e.g., [5, p. 130], [14, Exercise 2.2.23]). However, we can easily prove the general case in a direct manner by a much simplified version of the construction for Theorem 1.

Proposition 1 *For every continuous computable measure μ , and for every semi-computable discrete semimeasure P , there is a prefix-free machine T such that $P = Q_T^\mu$.*

Proof Let P be any semicomputable discrete semimeasure, with uniformly computable approximation functions f_t . We construct a prefix-free machine T in stages $s = \langle \sigma, t \rangle$. Let $D_s(\sigma) = \{\rho \in \mathbb{B}^* : (\rho, \sigma) \in T_s\}$. \square

Construction Let $T_0 = \emptyset$.

At stage $s = \langle \sigma, t \rangle$, if $\mu(\llbracket D_{s-1}(\sigma) \rrbracket) = f_t(\sigma)$ then let $T_s := T_{s-1}$.

Otherwise, let the set $R \subseteq \mathbb{B}^s$ of *available* strings be such that $\llbracket R \rrbracket = \mathbb{B}^\omega \setminus \llbracket \bigcup_{\tau \in \mathbb{B}^*} D_{s-1}(\tau) \rrbracket$. Collect in the auxiliary set A_s a number of available strings ρ from R with $\sum_{\rho \in A_s} \mu(\llbracket \rho \rrbracket)$ maximal but bounded by $f_t(\sigma) - \mu(\llbracket D_{s-1}(\sigma) \rrbracket)$, the amount of measure the current descriptions fall short of the latest approximation of $P(\sigma)$. Put $T_s := T_{s-1} \cup \{(\rho, \sigma) : \rho \in A_s\}$.

Verification It is immediate from the construction that $\bigcup_{\sigma \in \mathbb{B}^*} D_s(\sigma)$ is prefix-free at all stages s , so $T = \lim_{s \rightarrow \infty} T_s$ is a prefix-free machine. To show that $Q_T^\mu(\sigma) = \lim_{s \rightarrow \infty} \mu(\llbracket D_s(\sigma) \rrbracket)$ equals $P(\sigma)$ for all $\sigma \in \mathbb{B}^*$, it suffices to demonstrate that for every $\delta > 0$ there is some stage s_0 where $\mu(\llbracket D_{s_0}(\sigma) \rrbracket) > P(\sigma) - \delta$.

Choose positive $\delta' < \delta$. Wait for a stage $s_0 = \langle \sigma, t_0 \rangle$ with $\mu(\llbracket \rho \rrbracket) \leq \delta - \delta'$ for all $\rho \in \mathbb{B}^{s_0}$ and $f_{t_0}(\sigma) > P(\sigma) - \delta'$. Clearly, the available μ -measure

$$\begin{aligned} \mu(\llbracket R \rrbracket) &= 1 - \sum_{\tau \in \mathbb{B}^*} \mu(\llbracket D_{s_0-1}(\tau) \rrbracket) \\ &\geq 1 - \mu(\llbracket D_{s_0-1}(\sigma) \rrbracket) - \sum_{\tau \in \mathbb{B}^* \setminus \{\sigma\}} P(\tau) \\ &\geq P(\sigma) - \mu(\llbracket D_{s_0-1}(\sigma) \rrbracket) \\ &\geq f_{t_0}(\sigma) - \mu(\llbracket D_{s_0-1}(\sigma) \rrbracket). \end{aligned}$$

Consequently, if not already $\mu(\llbracket D_{s_0-1}(\sigma) \rrbracket) > P(\sigma) - \delta$, then the construction collects in A_{s_0} a number of descriptions of length s_0 from R such that $\mu(\llbracket D_{s_0}(\sigma) \rrbracket) = \mu(\llbracket D_{s_0-1}(\sigma) \rrbracket) + \sum_{\rho \in A_{s_0}} \mu(\llbracket \rho \rrbracket) > P(\sigma) - \delta$ as required.

Corollary 2 *For every continuous computable measure μ ,*

$$\{Q_T^\mu\}_T = \mathcal{P}.$$

2 The A Priori Semimeasures

In this section we show that the class of a priori semimeasures can be characterized as the class of universal transformations of any continuous computable measure. Section 2.1 introduces the class of a priori semimeasures. Section 2.2 is an interlude devoted to the representation of the a priori semimeasures as *universal mixtures*.

Section 2.3 presents the generalized characterization, and concludes with a brief discussion of how this reflects on the association with foundational principles.

2.1 A Priori Semimeasures

2.1.1 Universal Machines

Let $\{\rho_e\}_{e \in \mathbb{N}} \subseteq \mathbb{B}^*$ be any computable prefix-free and non-repeating enumeration of finite strings, that will serve as an encoding of some computable enumeration $\{M_e\}_{e \in \mathbb{N}}$ of all monotone machines. We say that a monotone machine U is *universal (by adjunction)* if for some such encoding $\{\rho_e\}_{e \in \mathbb{N}}$, we have for all $\rho, \sigma \in \mathbb{B}^*$ that

$$(\rho_e \rho, \sigma) \in U \Leftrightarrow (\rho, \sigma) \in M_e.$$

By a universal machine we will mean a monotone machine that is universal by adjunction. Contrast this to *weak universality*, which is the more general property that for all M there is a $c_M \in \mathbb{N}$ such that

$$(\rho, \sigma) \in M \Rightarrow \exists \rho' (|\rho'| < |\rho| + c_M \ \& \ (\rho, \sigma) \in U).$$

2.1.2 A Priori Semimeasures

We call a transformation by a universal machine a *universal transformation*. The a priori semimeasures are the universal transformations of the uniform measure.

Definition 7 An *a priori semimeasure* is defined by

$$\lambda_U(\sigma) := \lambda(\llbracket \{\rho : \exists \sigma' \succ \sigma ((\rho, \sigma') \in U) \} \rrbracket)$$

for universal monotone machine U .

Let \mathcal{A} denote the class $\{\lambda_U\}_U$ of a priori semimeasures. The next result implies that every element of \mathcal{A} can also be obtained as the transformation of λ by a machine that is *not* universal.

Proposition 2 For every continuous computable measure μ , there is for every semi-computable semimeasure ν a non-universal monotone machine M such that $\nu = \mu_M$.

Proof Let U be an arbitrary universal machine. We will adapt the construction of Theorem 1 of a machine M with $\mu_M = \nu$ in such a way that for every constant $c \in \mathbb{N}$ there is a σ such that for some ρ' with $(\rho', \sigma) \in U$, we have that $|\rho| > |\rho'| + c$ for all ρ with $(\rho, \sigma) \in M$. This ensures that M is not even weakly universal. \square

Construction The only change to the earlier construction is that at stage s we try to collect available strings of length l_s , where l_s is defined as follows. Let $l_0 = 0$. For $s = \langle \sigma, t \rangle$ with $t > 0$, let $l_s = l_{s-1} + 1$. In case $s = \langle \sigma, 0 \rangle$, enumerate pairs in U until a pair (ρ', σ) for some ρ' is found. Let $l_s := \max\{l_{s-1} + 1, |\rho'| + s\}$.

Verification The verification that $\mu_M = \nu$ proceeds as before. In addition, the construction guarantees that for every $c \in \mathbb{N}$, we have for σ with $c = \langle \sigma, 0 \rangle$ that $|\rho| > |\rho'| + c$ for the first enumerated ρ' with $(\rho', \sigma) \in U$ and all ρ with $(\rho, \sigma) \in M$.

We define a *discrete a priori semimeasure* in like manner.

Definition 8 A *discrete a priori semimeasure* is defined by

$$Q_U^\lambda(\sigma) := \lambda(\llbracket \{\rho : (\rho, \sigma) \in U\} \rrbracket)$$

for a *universal prefix-free machine* U , meaning that U is defined by

$$(\rho_e \rho, \sigma) \in U \Leftrightarrow (\rho, \sigma) \in T_e$$

for all $\rho, \sigma \in \mathbb{B}^*$ and some computable prefix-free and non-repeating enumeration $\{\rho_e\}_{e \in \mathbb{N}} \subseteq \mathbb{B}^*$ that serves as an encoding of some computable enumeration $\{T_e\}_{e \in \mathbb{N}}$ of all prefix-free machines.

2.2 Universal Mixtures

Every element of \mathcal{A} is equal to a *universal mixture*

$$\xi_W(\cdot) := \sum_{i \in \mathbb{N}} W(i) \nu_i(\cdot) \quad (2)$$

for some effective enumeration $\{\nu_i\}_{i \in \mathbb{N}} = \mathcal{M}$ of all semicomputable semimeasures, and some semicomputable *weight function* $W : \mathbb{N} \rightarrow [0, 1]$ that satisfies $\sum_{i \in \mathbb{N}} W(i) \leq 1$ and $W(i) > 0$ for all i . Conversely, one can show that every universal mixture equals λ_U for some universal machine U [22].

It is easy to see from the mixture form of the a priori semimeasures that every element of \mathcal{A} is *universal* in the sense that it *dominates* every other semicomputable semimeasure. That is, for every $\lambda_U \in \mathcal{A}$ there is for every $\nu \in \mathcal{M}$ a constant $c_\nu \in \mathbb{N}$, depending only on λ_U and ν , such that $\lambda_U(\sigma) \geq c_\nu^{-1} \nu(\sigma)$ for all $\sigma \in \mathbb{B}^*$. The converse does not hold: not all universal elements of \mathcal{M} are of the form λ_U or equivalently ξ_W . For instance, the sum of $\xi_W(\sigma)$ for all strings σ of the same length n will always fall short of 1 (because it does so for some semimeasures), but we can readily define a universal $\kappa \in \mathcal{M}$ with (say) $\kappa(\sigma) = \lambda(\sigma)$ for all σ up to a finite length n .

The aim of the current subsection is to strengthen the above statement of the equivalence of the a priori semimeasures and the universal mixtures, as follows.

First, let us call an enumeration $\{\nu_i\}_{i \in \mathbb{N}}$ of all semicomputable semimeasures *acceptable* if it is generated from an enumeration $\{M_i\}_i$ of all monotone Turing machines by the procedure of Theorem 1, i.e., $\nu_i = \lambda_{M_i}$. This terminology matches that of the definition of *acceptable numberings* of the partial computable functions [18, p. 41]. Every effective listing of all Turing machines yields an acceptable numbering. Importantly, any two acceptable numberings differ only by a computable permutation [17]; in our case, for any two acceptable enumerations $\{\nu_i\}_i$ and $\{\bar{\nu}_i\}_i$ there is a computable permutation $f : \mathbb{N} \rightarrow \mathbb{N}$ of indices such that $\bar{\nu}_i = \nu_{f(i)}$.

Furthermore, let us call a semicomputable weight function W *proper* if $\sum_i W(i) = 1$; this implies that W is computable.

Then we can show that for any acceptable enumeration of all semicomputable semimeasures, all elements in \mathcal{A} are expressible as some mixture with a *proper* weight function over *this* enumeration.

Proposition 3 *For every acceptable enumeration $\{v_i\}_i$ of \mathcal{M} , every element in \mathcal{A} is equal to $\xi_W(\cdot) = \sum_i W(i)v_i(\cdot)$ for some proper W .*

Proof Given $\lambda_U \in \mathcal{A}$, with enumeration $\{M_i\}_i$ of all monotone machines corresponding to U . We know that λ_U is equal to $\bar{\xi}_{\bar{W}}(\cdot) = \sum_i \bar{W}(i)\bar{v}_i(\cdot)$ for some acceptable enumeration $\{\bar{v}_i\}_i = \{\lambda_{M_i}\}_i$ of \mathcal{M} and semicomputable weight function \bar{W} . First we show that $\bar{\xi}_{\bar{W}}$ is equal to $\xi_{W'}(\cdot) = \sum_i W'(i)v_i(\cdot)$ for given acceptable enumeration $\{v_i\}_i$ and a semicomputable W' ; then we show that it is also equal to $\xi_W(\cdot) = \sum_i W(i)v_i(\cdot)$ for proper W .

Since enumerations $\{v_i\}_i$ and $\{\bar{v}_e\}_e$ are both acceptable, there is a 1-1 computable f such that $\bar{v}_i = v_{f(i)}$. Then

$$\begin{aligned} \sum_i \bar{W}(i)\bar{v}_i(\cdot) &= \sum_i \bar{W}(i)v_{f(i)}(\cdot) \\ &= \sum_i \bar{W}(f^{-1}(i))v_i(\cdot) \\ &= \sum_i W'(i)v_i(\cdot), \end{aligned}$$

with $W' : i \mapsto \bar{W}(f^{-1}(i))$.

We proceed with the description of a proper W . The idea is to have W assign to each i a positive computable weight that does not exceed $W'(i)$, additional computable weight to the index of a single suitably defined semimeasure in order to regain the original mixture, and all of the remaining weight to an “empty” semimeasure.

Let $q \in \mathbb{Q}$ be such that $\xi_{W'}(\epsilon) < q < 1$, and let c be such that $\sum_i 2^{-i-c} < 1 - q$. Let $W'_0(i)$ denote the first approximation of semicomputable $W'(i)$ that is positive. We now define computable $g : \mathbb{N} \rightarrow \mathbb{Q}$ by

$$g(i) = \min\{2^{-i-c}, W'_0(i)\}.$$

Clearly, $\sum_i g(i) < 1 - q$. Moreover, $\sum_i g(i)$ is computable because for any $\delta > 0$ we have a $j \in \mathbb{N}$ with $\sum_{i>j} 2^{-i-c} < \delta$, hence $\sum_{i \leq j} g(i) < \sum_i g(i) < \sum_{i \leq j} g(i) + \delta$.

Next, define $\pi(\cdot) = q^{-1} \sum_i (W'(i) - g(i)) v_i(\cdot)$. This is a semimeasure because $\pi(\epsilon) \leq q^{-1} \xi_{W'}(\epsilon) < q^{-1} q = 1$. Let k be such that $v_k = \pi$, and let l be such that v_l is the “empty” semimeasure with $v(\sigma) = 0$ for all $\sigma \in \mathbb{B}^*$ (both indices exist even if we cannot effectively find them).

Finally, we define W by

$$W(i) = \begin{cases} g(i) & \text{if } i \neq k, l \\ g(i) + q & \text{if } i = k \\ 1 - q - \sum_{j \neq l} g(j) & \text{if } i = l \end{cases}.$$

Weight function W is computable and indeed proper, and

$$\begin{aligned} \sum_i W(i) v_i(\cdot) &= \sum_i g(i) v_i(\cdot) + q v_k(\cdot) + 0 \\ &= \sum_i g(i) v_i(\cdot) + \sum_i (W'(i) - g(i)) v_i(\cdot) \\ &= \sum_i W'(i) v_i(\cdot). \end{aligned}$$

□

As a kind of converse, we can derive that any universal mixture is also equal to a universal mixture with a *universal* weight function, i.e., a weight function W such that for all other W' there is a $c_{W'}$ with $W(i) \geq c_{W'}^{-1} W'(i)$ for all i .

Proposition 4 *For every acceptable enumeration $\{v_i\}_i$ of \mathcal{M} , every element in \mathcal{A} is equal to $\xi_W(\cdot) = \sum_i W(i) v_i(\cdot)$ for some universal W .*

Proof By the above proposition we know that any given element in \mathcal{A} equals $\xi_{W'}(\cdot) = \sum_i W'(i) v_i(\cdot)$ for some (computable) W' over given $\{v_i\}_i$. Let k be such that $v_k(\cdot) = \sum_i 2^{-K(i)} v_i(\cdot)$, with $K(i)$ the prefix-free Kolmogorov complexity (via some universal prefix-free machine U) of the i -th lexicographically ordered string; $2^{-K(\cdot)}$ is a universal weight function. Define

$$W(i) = \begin{cases} W'(i) + W'(k) \cdot 2^{-K(i)} & \text{if } i \neq k \\ W'(k) \cdot 2^{-K(i)} & \text{if } i = k \end{cases},$$

which is a weight function because $\sum_i W(i) < \sum_{i \neq k} W'(i) + W'(k) = \sum_i W'(i)$. Moreover, W is universal because $2^{-K(\cdot)}$ is, and

$$\begin{aligned} \sum_i W(i) v_i(\cdot) &= \sum_{i \neq k} W'(i) v_i(\cdot) + W'(k) \sum_i 2^{-K(i)} v_i(\cdot) \\ &= \sum_i W'(i) v_i(\cdot). \end{aligned}$$

□

Hutter [8, p. 102–03] argues that a universal mixture with weight function $2^{-K(i)}$ is *optimal* among all universal mixtures, essentially because this weight function is universal. The above result shows that this optimality is meaningless: *every* universal mixture can be represented so as to have a universal weight function.

2.3 The Generalized Characterization

We are now ready to show that the universal transformations of any continuous computable measure μ yield the same class \mathcal{A} of a priori semimeasures. A minor caveat is that we will need to restrict the universal machines U to those machines with associated encodings $\{\rho_e\}_e$ that do not receive measure 0 from μ : so $\mu(\llbracket \rho_e \rrbracket) > 0$ for all $e \in \mathbb{N}$. Call (the associated encodings of) those machines *compatible* with measure μ . This is clearly no restriction for measures that give positive probability to every finite string (such as the uniform measure): all machines are compatible with such measures.

We will prove:

Theorem 2 *Let $\mu, \bar{\mu}$ be continuous computable measures. For any universal machine U that is compatible with μ , there is a universal machine V such that $\mu_U = \bar{\mu}_V$.*

It follows that $\{\mu_U\}_U = \{\bar{\mu}_V\}_V$ for any two continuous computable μ and $\bar{\mu}$, with U ranging over those universal machines compatible with μ and V over those universal machines compatible with $\bar{\mu}$. In particular, since λ is itself a continuous computable measure, we have that $\{\mu_U\}_U = \mathcal{A}$.

Our proof strategy is to expand the approach taken in [22] to show the coincidence of the a priori semimeasures and the universal mixtures. Let us first derive the fact that a universal transformation of μ is an a priori semimeasure.

Proposition 5 *Let μ be a continuous computable measure and let U be a universal machine compatible with μ . Then $\mu_U \in \mathcal{A}$.*

The proof rests on a fixed-point lemma that is a refined version of Corollary 1. For given encoding $\{\rho_e\}_e$, define $\mu^{\rho_e}(\cdot) := \mu(\cdot \mid \llbracket \rho_e \rrbracket)$ for any $e \in \mathbb{N}$. Here the conditional measure $\mu(\llbracket \tau \rrbracket \mid \llbracket \sigma \rrbracket) := \frac{\mu(\llbracket \sigma \tau \rrbracket)}{\mu(\llbracket \sigma \rrbracket)}$ for any $\sigma, \tau \in \mathbb{B}^*$. Let μ_M^ρ denote the transformation of μ^ρ by M .

Lemma 1 *Given encoding $\{\rho_e\}_{e \in \mathbb{N}}$ of the monotone machines as above. For every continuous computable measure μ ,*

$$\{\mu_{M_e}^{\rho_e}\}_e = \mathcal{M}.$$

Proof Let ν be any semicomputable semimeasure. Since μ^{ρ_e} is obviously a continuous computable measure for every $e \in \mathbb{N}$, by the construction of Theorem 1 we obtain for every e a monotone machine M with $\nu = \mu_M^{\rho_e}$. Indeed, there is a total computable function $g : \mathbb{N} \rightarrow \mathbb{N}$ that for given e retrieves an index $g(e)$ in the given enumeration $\{M_e\}_{e \in \mathbb{N}}$ such that $\nu = \mu_{M_{g(e)}}^{\rho_e}$. But by Kleene's Recursion Theorem [18], there must be a fixed point \hat{e} such that $M_{g(\hat{e})} = M_{\hat{e}}$, hence $\mu_{M_{\hat{e}}}^{\rho_{\hat{e}}} = \mu_{M_{g(\hat{e})}}^{\rho_{\hat{e}}}$.

This shows that for every ν there is an index e such that $\nu = \mu_{M_e}^{\rho_e}$. Conversely, the function $\mu_{M_e}^{\rho_e}$ is a semicomputable semimeasure for every e . \square

Proof of Proposition 5 Given continuous computable μ and universal U compatible with μ . We write out

$$\begin{aligned}\mu_U(\sigma) &= \mu(\llbracket \{\rho : \exists \sigma' \succ \sigma ((\rho, \sigma') \in U)\} \rrbracket) \\ &= \sum_e \mu(\llbracket \{\rho_e \rho : \exists \sigma' \succ \sigma ((\rho, \sigma') \in M_e)\} \rrbracket) \\ &= \sum_e \mu(\llbracket \rho_e \rrbracket) \mu(\llbracket \{\rho : \exists \sigma' \succ \sigma ((\rho, \sigma') \in M_e)\} \rrbracket \mid \llbracket \rho_e \rrbracket) \\ &= \sum_e \mu(\llbracket \rho_e \rrbracket) \mu_{M_e}^{\rho_e}(\sigma).\end{aligned}$$

Lemma 1 tells us that the $\mu_{M_e}^{\rho_e}$ range over all elements in \mathcal{M} . Moreover, $W(e) := \mu(\llbracket \rho_e \rrbracket)$ is a weight function because $\{\rho_e\}_e$ is prefix-free and U is compatible with μ , so μ_U is a universal mixture. \square

We now proceed to prove that every universal transformation of μ indeed equals some universal transformation of $\bar{\mu}$.

Proof of Theorem 2 Given continuous computable μ and $\bar{\mu}$, and universal U compatible with μ . Write out as before

$$\mu_U(\sigma) = \sum_e \mu(\llbracket \rho_e \rrbracket) \mu_{M_e}^{\rho_e}(\sigma).$$

Note that the function

$$P(\sigma) = \begin{cases} \mu(\llbracket \sigma \rrbracket) & \text{if } \sigma = \rho_e \text{ for some } e \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}$$

is a semicomputable discrete semimeasure. Hence by Proposition 1 we can construct a prefix-free machine T that transforms $\bar{\mu}$ into P : so $Q_T^{\bar{\mu}} = P$. Denote $n_e := \#\{\tau : (\tau, \rho_e) \in T\}$ the number of T -descriptions of ρ_e , and let $\langle \cdot, \cdot \rangle : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{N}$ be a partial computable pairing function that maps the pairs (e, i) with $i < n_e$ onto \mathbb{N} . Let $\tau_{\langle e, i \rangle}$ be the i -th enumerated T -description of ρ_e . We then have

$$\begin{aligned}\sum_e \mu(\llbracket \rho_e \rrbracket) \mu_{M_e}^{\rho_e}(\sigma) &= \sum_e Q_T^{\bar{\mu}}(\rho_e) \mu_{M_e}^{\rho_e}(\sigma) \\ &= \sum_e \sum_{i < n_e} \bar{\mu}(\llbracket \tau_{\langle e, i \rangle} \rrbracket) \mu_{M_e}^{\rho_e}(\sigma).\end{aligned}$$

Now for every $\langle e, i \rangle$ for which $\tau_{\langle e, i \rangle}$ becomes defined we can run the construction of Theorem 1 on $\bar{\mu}^{\tau_{\langle e, i \rangle}}$ and $\mu_{M_e}^{\rho_e}$. In this way we obtain an enumeration of machines $\{N_d\}_d$ such that $\bar{\mu}_{N_{\langle e, i \rangle}}^{\tau_{\langle e, i \rangle}} = \mu_{M_e}^{\rho_e}$ (with $i < n_e$) for all e . Then

$$\sum_e \sum_{i < n_e} \bar{\mu}(\llbracket \tau_{\langle e, i \rangle} \rrbracket) \mu_{M_e}^{\rho_e}(\sigma) = \sum_d \bar{\mu}(\llbracket \tau_d \rrbracket) \bar{\mu}_{N_d}^{\tau_d}(\sigma),$$

which we can rewrite to $\bar{\mu}_V(\sigma)$, defining V by $(\tau_d \rho, \sigma) \in V \Leftrightarrow (\rho, \sigma) \in N_d$.

It remains to verify that V is in fact universal. Namely, we cannot take for granted that $\{N_d\}_d$ is an enumeration of *all* machines, whence it is not clear that V is universal.² Note that it is enough if there were a single universal machine V' in $\{N_d\}_{d \in \mathbb{N}}$, but even that is not obvious (by Proposition 2 we know that for all continuous computable μ there are for any universal U non-universal N such that $\mu_N = \mu_U$).

However, there is a simple patch to the enumeration that guarantees this fact. Namely, given an arbitrary universal machine V' , we may simply put $N_d := V'$ at some $d = \langle e, i \rangle$ where it so happens that $\bar{\mu}_{V'}^{\tau_{(e,i)}} = \mu_{M_e}^{\rho_e}$. Our final objective is thus to show that $\bar{\mu}_{V'}^{\tau_{(e,i)}} = \mu_{M_e}^{\rho_e}$ for some e, i . To that end, define computable function $g : \mathbb{N} \rightarrow \mathbb{N}$ by $\mu_{M_{g(e)}}^{\rho_e} = \bar{\mu}_{V'}^{\tau_{(e,0)}}$. Since $Q_T^{\bar{\mu}}(\rho_e) > 0$ for each e , the string $\tau_{(e,0)}$ is defined for each e . Hence $\bar{\mu}_{V'}^{\tau_{(e,0)}}$ is defined, and g , that retrieves the index $g(e)$ of a machine that transforms μ^{ρ_e} to this semimeasure, is total. Then by the Recursion Theorem there is an index \hat{e} such that $M_{\hat{e}} = M_{g(\hat{e})}$, so $\mu_{M_{\hat{e}}}^{\rho_{\hat{e}}} = \mu_{M_{g(\hat{e})}}^{\rho_{\hat{e}}} = \bar{\mu}_{V'}^{\tau_{(\hat{e},0)}}$. \square

Corollary 3 *For any continuous computable μ , and U ranging over those universal machines that are compatible with μ ,*

$$\{\mu_U\}_U = \mathcal{A}.$$

Discrete versions of the above results are derived in an identical manner. Ultimately, we have the following discrete analogue to Corollary 3, where we let \mathcal{Q} denote the class of all discrete a priori semimeasures.

Proposition 6 *For any continuous computable μ , and U ranging over those prefix-free machines that are compatible with μ ,*

$$\{Q_U^\mu\}_U = \mathcal{Q}.$$

2.3.1 Discussion

We now return to the association of the function λ_U (as well as its discrete counterpart Q_U^λ) with foundational principles.

First, there is the association with the principle of *insufficient reason* or *indifference*. This is the principle that in the absence of discriminating evidence, probability should be equally distributed over all possibilities. Solomonoff writes, “If we consider the input sequence to be the ‘cause’ of the observed output sequence, and we consider all input sequences of a given length to be equiprobable (since we have no a priori reason to prefer one rather than the other) then we obtain the present model of induction.” [20, p. 19]. Also see [12, 16].

Second, there is the association with Occam’s razor. Solomonoff writes, “That [this model] might be valid is suggested by ‘Occam’s razor,’ one interpretation of which is that the more ‘simple’ or ‘economical’ of several hypotheses is the more

²This is also an (overlooked) issue in the original proof [22, Lemma 4]. It is easily resolved by the same approach we will be taking here, where it is immediate that for given universal V there is an e with $\lambda_V = \nu_e$.

likely . . . —the most ‘simple’ hypothesis being that with the shortest ‘description.’” [20, p. 3]. Also see [2, 9, 13, 15, 21].

Note that so stated, these associations very much rely on the fact that the uniform measure λ always assigns larger probability to shorter strings, and equal probability to equal-length strings. This is a unique feature of λ . The results of this paper, however, show that the choice of the uniform measure in defining algorithmic probability is only circumstantial: we could pick *any* continuous computable measure, and still obtain, as the universal transformations of *this* measure instead of λ , the very same class of a priori semimeasures. This suggests that properties derived from the presence of λ in the definition are artifacts of a particular choice of characterization rather than an indicative property of algorithmic probability, and hence undermines both associations insofar as they indeed hinge on the uniform measure.

Acknowledgements This research was supported by NWO Vici project 639.073.904. I am grateful to the anonymous reviewers for their thoughtful comments, to Alexander Shen for valuable remarks on an earlier version of this paper, to Peter Grünwald, Jan Leike, and Daniël Noom for helpful discussions, and to Jeanne Peijnenburg for the question that initiated this work.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Chaitin, G.J.: A theory of program size formally identical to information theory. *J. Assoc. Comput. Mach.* **22**(3), 329–340 (1975)
2. Cover, T.M., Thomas, J.A.: *Elements of Information Theory*, 2nd edn. Wiley, Hoboken (2006)
3. Day, A.R.: On the computational power of random strings. *Annals of Pure and Applied Logic* **160**, 214–228 (2009)
4. Day, A.R.: Increasing the gap between descriptive complexity and algorithmic probability. *Trans. Am. Math. Soc.* **363**(10), 5577–5604 (2011)
5. Downey, R.G., Hirschfeldt, D.R.: *Algorithmic randomness and complexity*. Springer, New York (2010)
6. Gács, P.: On the symmetry of algorithmic information. *Soviet Mathematics Doklady* **15**(5), 1477–1480 (1974)
7. Gács, P.: Expanded and improved proof of the relation between description complexity and algorithmic probability. Unpublished manuscript (2016)
8. Hutter, M.: *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability*. Springer, Berlin (2005)
9. Hutter, M.: On universal prediction and Bayesian confirmation. *Theor. Comput. Sci.* **384**(1), 33–48 (2007)
10. Levin, L.A.: On the notion of a random sequence. *Soviet Mathematics Doklady* **14**(5), 1413–1416 (1973)
11. Levin, L.A.: Laws of information conservation (nongrowth) and aspects of the foundation of probability theory. *Probl Inf Transm* **10**(3), 206–210 (1974)
12. Li, M., Vitányi, P.M.B.: Philosophical issues in Kolmogorov complexity. In: Kuich, W. (ed.) *Proceedings of the 19th International Colloquium on Automata, Languages and Programming*, pp. 1–16. Springer (1992)
13. Li, M., Vitányi, P.M.B.: *An Introduction to Kolmogorov Complexity and Its Applications*, 3rd edn. Springer, New York (2008)

14. Nies, A.: Computability and randomness. Oxford University Press (2009)
15. Ortner, R., Leitgeb, H.: Mechanizing induction. In: Gabbay, D.M., Hartmann, S., Woods, J. (eds.) *Inductive Logic*, volume 10 of *Handbook of the History of Logic*, pp. 719–772. Elsevier (2011)
16. Rathmanner, S., Hutter, M.: A philosophical treatise of universal induction. *Entropy* **13**(6), 1076–1136 (2011)
17. Rogers, H., Jr.: Gödel numberings of partial recursive functions. *J. Symb. Log.* **23**(3), 331–341 (1958)
18. Rogers, H., Jr.: *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, New York (1967)
19. Schnorr, C.-P.: Process complexity and effective random tests. *J. Comput. Syst. Sci.* **7**, 376–388 (1973)
20. Solomonoff, R.J.: A formal theory of inductive inference. Parts I and II. *Inf Control* **7**(1–22), 224–254 (1964)
21. Solomonoff, R.J.: The discovery of algorithmic probability. *J. Comput. Syst. Sci.* **55**(1), 73–88 (1997)
22. Wood, I., Sunehag, P., Hutter, M.: (Non-)equivalence of universal priors. In: Dowe, D.L. (ed.) *Papers from the Solomonoff Memorial Conference, Lecture Notes in Artificial Intelligence 7070*, pp. 417–425. Springer (2013)
23. Zvonkin, A.K., Levin, L.A.: The complexity of finite objects and the development of the concepts of information and randomness by means of the theory of algorithms. *Russ. Math. Surv.* **26**(6), 83–124 (1970)